面向图书馆资源建设的国外开放科技资源项目实证研究及启示

Empirical Study on Foreign Open Resources Project Oriented to Open Collection Development of Library

陈雪飞 黄金霞 张建勇 黄永文 王昉

（中国科学院国家科学图书馆 北京 100190）

[摘要] 开展开放科技信息资源建设对促进科技信息资源的传播和利用，提升图书馆信息服务能力有重大意义。选取当前 5 个国外著名的开放科技资源建设项目作为研究样本，对样本项目从资源发现及获取、质量控制、利用机制、服务开发 4 个方面进行了考察。在借鉴这些国外优秀项目经验基础上，我国图书馆在开展开放资源建设时，要注重资源评价遴选、规范资源利用机制、重视资源长期保存工作。

[关键词] 开放科技信息资源　资源建设　实证研究　开放资源项目

[中图分类号]G250

[文献标识码]A

[Abstract] Open academic and technology collection development is of great significance in promoting the dissemination and exploitation of open academic resources and the service capabilities of libraries. The empirical investigation showed the development status of 5 foreign open academic resources projects from resource discovery mechanism, quality control, utilization rules, service design. Especially, at the basis of the demonstration experience of these outstanding projects, We should continue to improve and pay attention to evaluation and selection, utilization and long-term preservation in the open collection development work.

[Keyword] Open Academic and Technology Resources; Collection Development; Empirical Study; Project

开放科技资源是当前 e-science 环境下出现的一种新型学术交流和知识传播模式，已经成为科研人员获取知识的重要来源。张晓林[1-2]等提出，开放资源正逐步逼近“成为学术研究主流资源”的转折点，这种趋势对习惯于依赖采购资源来建设馆藏、利用馆藏提供服务的图书馆形成严峻挑战。同时，在商业资源采购成本不断上涨和用户资源获取方式多元化的压力下，“抓住”开放资源，也成为图书馆为用户提供更好服务的发展契机。

在开放大趋势下，国际上一些图书馆、科研机构、科研资助机构、数据库商等陆续开展了开放科技资源项目，很多优秀的开放科技资源已经受到科研人员的广泛认可。为了使我国的科研人员能够稳定、持续、迅速地获取和利用开放科技资源，我们必须充分了解和分析开放资源的建设模式，并有选择地将开放资源纳入我国图书馆的信息资源体系。本文在界定开放科技资源范畴的基础上，选取国外若干开放科技资源项目，对这些项目的资源建设机制进行考察，以期为我国图书馆开展开放科技资源建设工作提供借鉴。

1 开放科技资源建设范畴

开放科技资源由 OA 资源发展而来，OA 资源是指开放存取的资源，即通过“绿色之路”和“金色之路”实现开放的科技信息资源。然而，面向科研信息需求的开放科技资源范畴不应局限于 OA 资源。黄如花[3]等人提出了广义的 OA 资源概念，即任何用户均可免费在线获取的、不受许可限制的所有数字化学术信息资源。SPARC[4]也对资源的开放性进行了定义，认为任何可免费使用、复用、再传播的内容或数据都属于开放资源。从图书馆资源建设角度考虑，本文采用了 SPARC 的定义。

开放科技资源涵盖常规科技信息资源和非常规科技信息资源的所有内容类型，如论文后印本、预印本、正式出版著作、未正式出版手稿、项目网站、会议论文、学位论文、学术blog、专利、数据、实验记录、标准、程序、教学资料、图表、甚至学术交流网络等。对这些科技资源进行发现获取、评价遴选、组织揭示、利用、保存以及开发第三方开放服务等都属于开放科技资源建设的范畴。

2 开放科技资源项目实证分析

从资源内容的揭示层次看，开放科技资源项目分元数据集成模式和全文集成模式两种，元数据集成模式又包括载体元数据集成和内容元数据集成两种类型，全文集成模式包括学科仓储、机构仓储、OA 出版三种类型。本文从每种类型选取了一个代表性项目作为分析样本，选择标准包括项目的学术知名度、建设单位类型、使用量、资源量、运行机制的健全性等。表 1 即为 5 个示范项目的部分特征：

表 1　5 个开放科技资源项目的基本特征

项目名称	集成类型	资源类型	建设单位	语种	资源量
OPENDOAR[5]	载体元数据集成	开放仓储目录	瑞典伦德大学	英语界面；多语种内容	>2200 个仓储
BASE[6]	内容元数据集成	开放资源搜索引擎	德国比勒菲尔德大学	英语、德语界面；22 种欧洲语言内容	>2 亿篇文献；1200 种学术资源
PMC[7]	学科仓储	学科仓储	美国 NCBI	英语	>240 万篇论文
eDoC[8]	机构仓储	机构仓储	德国马普学会	英语；德语	未知
PLOS[9]	OA 出版	开放期刊	PLOS	英语	7 种期刊

开放科技资源建设是对网络上散落的开放科技信息资源进行发现、获取、遴选、提供利用服务的过程，与商业信息资源建设过程有着很大差异。本文以表中 5 个代表性项目作为示范样本，从资源发现获取机制、质量控制机制、开放利用机制和开放服务方式 4 个方面进行分析，从中探讨可用于指导我国开放科技资源建设的有益模式。

2.1 资源发现及获取机制

资源发现和获取是开放科技资源建设的基础，直接影响资源的数量、质量以及可持续利用。资源发现获取的方式包括主动和被动两种方式，主动方式是指资源内容采用人工采集、特定检索协议、爬虫抓取等方式建设，被动方式指内容由用户或第三方提供，可以是资源所有者主动登记/提交、自由推荐、会员机构或出版商批量贡献等。

元数据集成模式的 2 个项目中，OPENDOAR 主要采取被动方式，由推荐者注册登记资源，成为全球 OAR 的资源建设中心[10]。但覆盖率并不理想，据笔者对比发现，OPENDOAR 和 Ranking web of repositories[11]收录的知识仓储互不覆盖率接近 30%。另外，在资源更新性和连续性方面也存在不足， 如 OPENDOAR 并没有覆盖中国科学院许多新建的机构知识库，收录的中国科学院机构仓储只占中科院所有机构仓储的 1/4 左右（25 个）。内容元数据集成模式一般都是架构在已有开放资源集合上的检索服务，通过检索协议或爬虫等技术方式发现和获取资源。但开放学术资源的搜索引擎还处在发展初期阶段[12]，BASE 是目前开放学术搜索引擎中较为成熟的典型，目前覆盖了 2417 个机构知识库，已经抓取了 1200 多种学术资源。它遵守 OAI-PMH，利用“收割机”收集文献资源，采用了 Norwegian 公司邻近搜索与传输技术，把密切相关数据库主题内容与搜索引擎直接搜索的相关资源组合在一起，实现了深网挖掘[13]。

对于全文集成模式来说，由于涉及到更复杂的使用权益问题，学术影响力或政策支持往往成为保障其合法获取资源的有效手段。以 PMC 为例，许多资助机构在开放存缴政策中都要求将受资助的论文存缴到 PMC 中，大多数论文都能在出版后一年内开放获取。eDOC 作为

德国马普学会的机构仓储，其运转依赖于本机构的存缴政策和经费支持。而 PLOS 等优秀的开放期刊由于其开放性使论文更易获取和传播，同时还邀请专家担任编委或投稿等，其学术影响力不断提高。据统计，2012 年 PLOS ONE 的影响因子为 4.092[14]，PLOS 无疑拥有了稳定并且优秀的稿源。下表总结了 5 个项目各自的资源获取方式及优缺点。

表 2　5 个开放科技资源项目的资源发现和获取方式及优缺点

项目名称	资源发现及获取方式	优点	缺点
OPENDOAR	用户推荐	用户参与度和积极性高	用户依赖性强，覆盖范围有限，持续发现和更新机制不完善。
BASE	机器抓取	资源量大，获取方便快速，新资源发现及时	发现资源并非全部为开放资源，用户参与度不够。
PMC	出版商、个人贡献	政策支持，资源发现持续稳定，保障性高	并不是完全的开放获取，大部分属于延迟开放。
eDoC	机构人员提交	政策支持，资源发现稳定持续，保障性高	资源有限单一，缺少跨库发现机制。
PLOS	学术吸引	形成资源获取良性循环机制	依赖于其学术影响力，具有一定风险。

2.2 质量控制机制

相比商业科技资源以同行评议作为质量控制的方法，开放科技资源的质量控制机制更具复杂性和多样性。首先，开放科技资源内容多样，既有期刊论文、图书、会议论文、学位论文等常规资源，也有博客、项目网站、软件、科学数据等非常规资源，即使同一种内容，也可能会有多种载体类型表示，需要采取灵活和多样化的评价、遴选机制；其次，开放科技资源的开放程度不一，利用方式不同，技术规则不统一，难以进行比较评价。

经考察发现，OPENDOAR 由专业人员对每家开放仓储按照学术水平、资源状况、运行情况等多方面进行评估和筛选[15]，采用了先纳入后审核的方法，一定程度保证了资源质量和开放性。但 OPENDOAR 并没有对资源进行进一步比较遴选，以向用户提供优质资源导航；BASE 作为一个搜索服务提供方， 没有提供明确的质量控制机制，只能依赖于资源提供方本身是否对资源有质量控制，暴露了搜索引擎通有的弊端。相比之下，全文集成类的开放科技资源质量控制标准较为严格。eDOC 从元数据、学术质量、格式等方面对提交文档进行多重审查，确保从内容到形式的规范化和高质量；PMC 提供了详细的期刊收录流程[16]（如图 1），从图中可以看出，为平衡资源数量和质量的关系，PMC 的质量要求可能会比 MEDLINE 略低，但 LSTRC 评议过程使这一差距变得很小。可以看出，对于这类有政策支持的开放科技资源项目，往往在质量控制上有更多优势和话语权，为制定完善的自有的质量标准提供了实现条件，可以作为领域参考规范或指南指导其他开放科技资源项目的资源评价遴选工作。

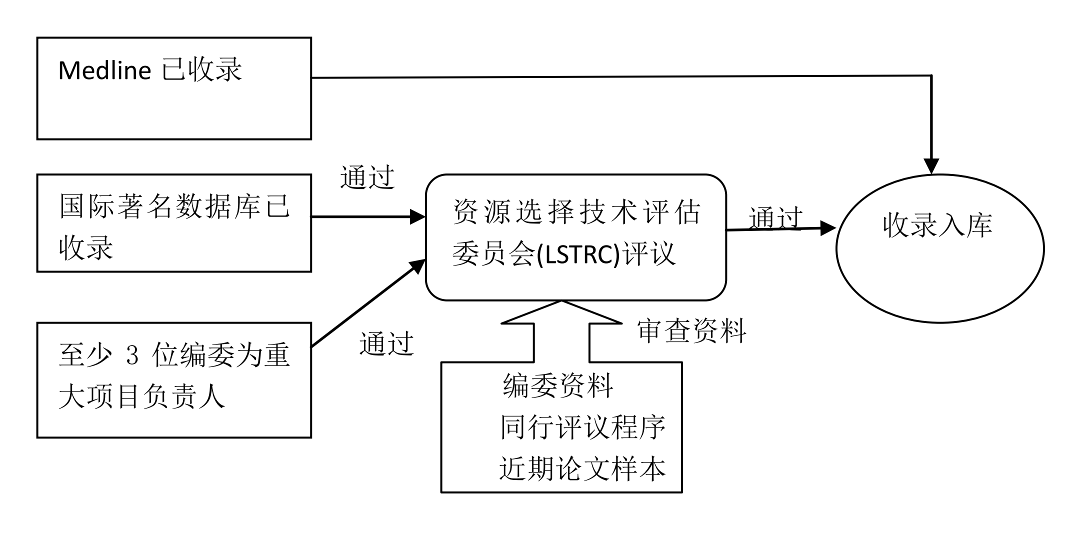


图 1 PMC 审议流程

2.3 开放利用机制

开放科技资源建设的根本目的是提供开放资源的利用，以最大限度地满足用户的信息需求。开放资源因其内容的开放性，在利用许可、知识产权、法律限定等方面存在较多约束，同时又缺乏统一规范，并且，由于不同系统的建设目标各异，各自关注的开放程度不同，这些都对开放资源的继续利用造成困难。本文通过分析 5 个开放科技资源项目的内容开放程度和使用许可政策，以期为开放科技资源的规范利用提供依据。

开放科技资源项目的访问权可分为 4 个等级：无任何限制开放、登陆注册开放、限制开放和不开放。这 5 个开放科技资源项目中，OPENDOAR、BASE、PMC、PLOS 都是无限制开放；eDOC 部功能有 IP 限制访问。但是，OPENDOAR、BASE 这两个元数据集成模式项目的无限制开放仅针对项目平台及其内容，收录的资源的开放性无法保证。如 OPENDOAR 收录的一些仓储中只有部分全文开放，而 BASE 大约只有 80%的索引文献开放，其他资源需要内容提供商授权才能获得。同时，对于一些全文集成模式的项目来说，也存在部分记录只有元数据没有全文的情况，因此，完全的开放在实际的开放科技资源建设中实现并不彻底。

目前，大部分成熟的项目一般采用许可协议作为使用权的规范方式，包括广为认同和采用的 CC 协议系列（cc-by-nc-nd，cc-by-nc-sa，cc-by-nc，cc-by-nd，cc-by-sa，cc-by），开源软件协议（GNU GPL ，GNU LGPL），开放数据协议，创新者专利协议等，以及地区性许可协议，如德国 DINI 认证（可作为 CC 协议的替代），Deutsche Free Software License 等。其中，OPENDOAR 采用了 cc-by-nc-sa 作为许可协议；BASE 是一个 OAI 服务提供者，是欧洲“DRIVER”项目的成员，因此也遵循“DRIVER”采用的 cc-by 协议； PMC 资源的版权属于作者和出版者，PMC 只提供存档，只有占资源少数的 Open Access Subset 使用 CC-by 授权；以 PLOS 为代表的 OA 出版物则对 CC 协议的普及率较高，采用意识较强。总体来看，尽管 CC 协议系列在开放资源中被广泛认同和采纳，但因为缺乏强制性和法律约束，许可协议的使用完全依赖资源建设者的开放意识和主动参与。

2.4 开放服务机制

开放服务有利用于促进开放获取成果的传播和利用，给用户带来好的使用体验，是开放科技资源走向成熟的重要标志。目前，开放科技资源项目开发的服务主要包括检索服务、情报服务（评价、计量统计、工具软件等）、API 服务、个性化服务（RSS、邮件、SNS 分享等）、知识链接服务等，下表列出了 5 个项目开通的服务类型。

表 3　5 个开放科技资源项目开通的服务类型

项目名称	服务
OPENDOAR	检索服务；统计服务；邮件服务；API 服务
BASE	检索服务；Blog / Twitter 服务；移动客户端服务；引文分析；搜索插

	件服务；Zotero 界面服务；API 服务
PMC	检索服务；跨库相关链接服务
eDOC	检索服务；跨库知识链接服务
PLOS	检索服务；博客、twitter、facebook、邮件订阅等个性化服务；

可以看出，5 个项目都开通了检索服务，OPENDOAR、BASE、PLOS 开通了个性化服务，说明资源建设者对资源的易用性和交互性比较重视，尤其是标签、博客、SNS 分享等应用也开始出现在平台上。但是情报服务普遍缺失，只有 OPENDOAR 开通了“repository Statistics”统计服务和政策工具服务，另外 BASE 也可以进行简单的引文分析。对于资源的增值开发和深度知识挖掘成为开放科技资源服务的盲点。知识链接服务方面，平台内资源的相互链接基本实现，但实现外部知识链接的较少，PMC 基于本身优势实现了跨库链接，可从参考文献直接链接到 PubMed、CrossRef，eDOC 也能实现外部资源动态链接，但对于机构外用户仅能实现对外部开放资源的链接。另外，有两个项目已知能提供 API 服务，支持第三方重用。可见，情报服务、知识互链服务、重用服务有较大的发展空间，可以更好地促进开放资源的深度利用。

3 对我国图书馆开放科技资源建设的启示

3.1 注重开放资源评价遴选

开放资源的评价遴选是开放资源建设中的关键环节之一，贯穿整个开放资源建设的发现、组织、利用、存储的全过程，关系资源体系的稳定和可靠利用，直接影响开放资源的服务效果和用户认同感。但开放资源评价和遴选的理论和方法缺失，没有既成的评价体系可供使用，并且从已有的项目实证分析来看，开放资源本身结构分散，特点各异，也造成对评价的无统一标准、操作性难的问题。参考已有的评价操作实践，笔者认为，在流程和制度规范方面，开放资源的评价遴选工作在资源发现阶段应分类型、分组织方式、分开放程度等建立评估制度，在评估基础上结合本机构用户需求进行资源遴选，建设初步的开放资源库，并在后期的组织集成、利用、保存、服务阶段实现动态质量监控。在评价方法和技术方面，设计一套开放式的、定量与定性方法相结合的评估指标体系，包括内容、技术、管理、利用许可、个性化指标、外界反馈（用户评价、图书馆员评价）等不同方面指标，每个指标下可再设置下一级指标，以进行同类资源的横向综合评估和不同方面评估。

3.2 完善开放资源利用机制

目前开放资源的利用机制并不完善，除了一部分成熟的项目采用了规范化的许可协议或在资源网站上申明使用权限外，互联网上大部分开放资源仍处于使用“无序化”状态，造成一定的知识产权风险，使用户在使用这些资源时心存疑虑，也阻碍了这些资源的进一步利用以支持再次知识创新。因此，我国图书馆在开放科技资源建设过程中要不断完善开放资源的利用机制。首先，资源建设者要在充分了解当前已有的开放资源使用许可政策、技术规则的基础上，设计开放资源的多种获取策略和利用方法，引导用户合理、便利地使用，规避可能的知识产权风险；其次，在遵守资源的使用规范基础上，开发可能的开放服务，如数据挖掘、情报服务、知识组织等，同时也推动图书馆的服务创新；再次，对于未有规范化使用机制的开放资源，图书馆一方面要将这些资源及时纳入资源体系中为用户提供“合理使用”范围的服务，一方面要以先驱者为己任，推动、倡导开放资源的使用规范化，提出开放资源的利用规范，与一些大型开放资源提供方进行谈判和合作以获取尽可能多的利用权限。

3.3 重视开放资源长期保存

随着开放资源逐步为科研人员所认知和使用，开放资源的长期保存也亟待提上日程。2010 年，中科院国家科学图书馆就与 BMC 签署了长期保存合作协议，实施了对 BMC 出版的 200 多种开放获取期刊的长期保存[17]，还有诸如 Internet Archive 的开放网络存档项目也对网络上零散的开放资源进行了规模保存。但大部分开放资源目前仍缺乏有效的长期保存机

制，在 OpenDOAR 所做的一项调查发现，大约有三分之二的 OA 仓储没有制定有关数字资源长期保存的政策[18]。因此，在开放资源建设中，开展资源的长期保存工作意义重大。建设开放科技资源的长期保存体系，应从权益管理和技术保障机制两方面加以探索，积极与开放资源提供方开展长期保存合作，在参考 BMC、DOAJ 等成熟的 OA 期刊长期保存案例基础上，积极开展多种类型资源的保存，如科研项目网站、博客、开放数据、开源科研软件、社会网络资源等，对这些资源一方面要积极争取摄取权、保存权、管理权，一方面也要有选择的进行保存，避免造成存储空间浪费。

参考文献

[1] 张晓林,李麟,刘细文,等.开放获取学术信息资源：逼近“主流化”转折点[J].图书情报工作,2012,56(9):42-47.

[2] 张晓林,曾燕,李麟,等.开放学术信息资源环境的挑战及其应对策略[J].图书情报工作,2012,56(19):76-84.

[3] 黄如花,刘贵玉.开放存取资源元数据管理的对策.[J].情报理论与实践,2009,32(10):5-8.

[4] Defining the Open in Open Data, Open Content and Open Services.[EB/OL].[2012-11-26]. http://opendefinition.org/

[5] OPENDOAR. [EB/OL].[2012-11-26].http://opendoar.org

[6] BASE .[EB/OL].[2012-11-26].http://www.base-search.net

[7] PMC .[EB/OL].[2012-11-26].http://www.ncbi.nlm.nih.gov/pmc/

[8] eDOC. [EB/OL].[2012-11-26].http://edoc.mpg.de/

[9] PLOS .[EB/OL].[2012-11-26].http://www.plos.org/

[10] 谭从容, OpenDOAR 的背景、特点和应用[J].图书馆建设,2008(1):49-51.

[11] [EB/OL].[2012-11-26].http://repositories.webometrics.info/

[12] 同[2].

[13] 赵金海.德国免费网络学术资源入口-比勒菲尔德学术搜索引擎（BASE）研究[J].中国索引,2009,7(3):7-12.

[14] JCR.[EB/OL]. [2012-02-20]. http://admin-apps.webofknowledge.com/JCR/JCR?PointOfEntry=Home&SID=1Dk9Bl1i4egPEo14Doc.

[15] Stevan Harnad. From ROAR to DOAR. [EB/OL].[2013-01-3].http://www.ecs.soton.ac.uk/~harnad/Hyper mail/Amsci/5095.htm.

[16] PMC Information for Publisher.[EB/OL].[2012-12-31].http://www.pubmedcentral.gov/about/pubinfo.Html.

[17] 张晓林.数字资源长期保存--责任、权利、战略、行动[R].2011.

[18] Peter Millington． Moving Forward with the OpenDOAR Directory,8th International Conference on Current Research Information Systems，Bergen，11th-13th May,2006 [EB/OL].[2012-11-25].http://www.opendoar.org/tools/en/policies.php.

作者简介

陈雪飞，女，1987 年生，硕士，助理馆员，发表论文 4 篇；
黄金霞，女，1972 年生，博士，副研究馆员，发表论文 10 余篇；
张建勇，男，1965 年生，研究馆员，资源建设部主任，发表论文 30 余篇；
黄永文，女，1976 年生，博士，副研究馆员，发表论文 20 余篇；
王昉，女，1979 年生，硕士，馆员，发表论文 10 余篇。